Ports, Roads, and Structural Change: The Role of Internal Geography in Trade Liberalization*

Anmol Agarwal
University of Virginia

Aishwarya Kekre
University of Virginia

November 27, 2024

Abstract

This paper examines the role of internal geography in shaping the spatial distribution of benefits from trade liberalization. Focusing on India's 1990s trade reforms, we show that high internal trade costs play a critical role in determining the response of local labor markets. Exploiting India's complete road network, we show that regions with better access to international markets via sea or air ports experience a stronger reallocation of labor from agriculture to manufacturing and services. This reallocation is driven by a higher growth in manufacturing output, labor productivity, and demand for manufacturing labor in regions with better international access. These regional disparities are quantitatively large and persist for up to a decade after trade liberalization.

*We thank Kerem Coşar, James Harrigan, John McLaren, Jonathan Colmer, and members of the Trade and Development Groups at UVa for their valuable comments. We are grateful to Treb Allen, Devaki Ghose, and Petia Topalova for sharing data which was instrumental for this project. All errors are our own.

1 Introduction

How does external integration affect local labor markets when internal integration is incomplete? Are regions better integrated with international markets more likely to experience stronger manufacturing gains following a large-scale trade liberalization or do they perish under greater foreign competition? Does better international access facilitate or hinder structural transformation? We address these fundamental questions by studying India's large-scale trade liberalization policy in the 1990s, which reversed decades of inward-looking economic policy. While external integration progressed rapidly, internal integration was far from complete with a poor road network characterized by slow average speeds and long delays. We provide detailed evidence that India's internal geography played a crucial role in shaping the spatial incidence of the effects of trade liberalization.

In the presence of high internal trade costs, regions better integrated with world markets can experience either stronger positive or negative effects of trade liberalization as compared to less integrated regions. Goldberg et al. (2010b) show that reduced import tariffs led to a massive surge in the import of intermediate inputs in India from 1991-97. The opportunity to take advantage of these cheaper and/or better imported inputs dissipates if the costs of transporting them from an international gate to the firm's location are high. Thus, firms in regions with better access to international markets are better positioned to gain from the spur in intermediate imports. Further, Topalova and Khandelwal (2011) show that a fall in both input and output tariffs led to a rise in firm productivity in the 90s in India. The lack of foreign goods in the domestic market, both intermediate and final, restricts the potential technology spillovers from superior products for more inland firms.

The negative effects stem from the fact that more inland firms are better protected from increased foreign competition due to a fall in output tariffs. With high internal trade costs, the effective price of foreign goods increases as they reach inland markets, thus, losing competitiveness. The local labor market effects are dependent on the magnitude and direction of these effects. If the positives outweigh the negatives, regions with better international access experience relatively larger manufacturing gains. These larger gains can translate into increased labor demand in manufacturing. Recent work has established that primary inputs and materials are complementary in production (Peter and Ruane, 2023; Castro-Vincenzi and Kleinman, 2022)[1]. A fall in tariffs thus reduces the price of intermediates and complementarity leads to an increase in spending share on labor. Ad-

[1] The study by Peter and Ruane (2023) is particularly well suited for our paper as the estimates are in the context of India and obtained using India's trade liberalization as an exogenous shock.

ditionally, a rise in TFP can lead to increased demand for all inputs, including labor. Better international access, therefore, directly facilitates structural transformation by increasing the opportunities of switching from agriculture to manufacturing[2].

In the absence of high internal costs, the above channels become less important as spatial disparities in access to international gates are less stark. For instance, a network of high-speed roads or rails connecting inland regions with international ports in an economy can dampen the initial location disadvantage. Until at least the early 2000s, India did not have such a network. Traveling a distance of 1,500 km by truck in India in 2000 took four to five days, which is about five times longer than in the United States (Asturias et al., 2019). Economic losses from congestion and poor roads are estimated to be as high as INR 120-300 billion ($\approx$ USD 2.6-6.5 billion) per year (Mohan et al., 1996). Major investments in highway upgrades only began in 2001 with the construction of the Golden Quadrilateral (GQ). The project was close to completion by 2006. Other major projects, such as the North-South and East-West corridors, had only 10% of the work completed by 2006. Therefore, India in the period from 1987 to 2004 provides an ideal setting to analyze how the effects of a major trade liberalization episode are spatially concentrated, particularly when access to international markets via ports is more costly for distant regions.

We combine several data sources to build a rich geo-coded dataset of firms in the manufacturing sector in India. We digitize the 1996 road map of India, published by the Government of India, to obtain accurate travel times. The digitized map allows us to identify five types of roads, each differing in quality. We assign average speeds along different roads consistent with the literature to compute the shortest travel time between a district's centroid and its nearest international gateway. From the computed travel times, we create a measure of the iceberg trade cost of moving goods between a district center and its nearest international gate. This serves as our measure of international market access i.e., lower trade cost means better international access. Leveraging the rich information from the firm-level data we create a district-industry panel extending to the pre-liberalization period. We show that in response to a decline in the effective rate of protection within an industry, the positives outweigh the negatives as the total output and labor productivity increase relatively more in districts with better access to international markets. We further show that both the total number of firms and total manufacturing employment increase relatively more in districts with better international access.

We provide evidence that access to international markets is crucial in assessing the response of local labor markets to new opportunities. Increase in manufacturing output

[2] The implications are reversed if the negative effects from increased export competition outweigh the positives.

and firms in more internationally integrated districts creates additional opportunities for workers in the manufacturing sector. It is well established that wages in the manufacturing sector in India are higher and less seasonal than in agriculture (ILO, 2018). Moreover, under the assumption of non-homothetic preferences, higher income further increases the demand for services. We use data from the Population Census and multiple rounds of nationally representative household surveys, to show that in response to trade liberalization, districts with a lower trade cost experience a larger reduction in the share of the labor force in agriculture and greater gains in the share of the labor force in manufacturing and services. Better international access fosters structural transformation in a period of falling external costs but high internal costs. The economic magnitude of these effects is large. Our benchmarking exercise shows that international access can explain, at least, an additional decline of 2 million workers in agriculture, and an additional gain of 0.5 million workers in manufacturing and 0.45 million workers in services in districts with below-median trade cost, relative to districts with above-median trade cost. Policymakers in developing countries often emphasize on structural transformation as an important development goal. Our results show that poor domestic transport infrastructure is an important determinant of winners and losers along this dimension. The implications extend well beyond the case of India as countries in the developing world, like Africa, continue to suffer from poor and inadequate road networks as they further complete their integration with world markets.

This paper relates to several strands of literature. The first examines the impact of internal trade costs on economic activity. Donaldson and Hornbeck (2016), Donaldson (2018), Baum-Snow et al. (2020), and Brooks and Donovan (2020) estimate the effects of market access, driven by changes in transport infrastructure, on regional economic outcomes. Specifically in the context of India, Van Leemput (2021) estimates that internal trade barriers make up 40% of total trade barriers in India. Several papers examine the gains from reducing internal trade barriers in India by examining road infrastructure improvements. While Asher and Novosad (2020) focus on rural roads, Ghani et al. (2016), Alder (2016), Asturias et al. (2019), and Vogel et al. (2024) examine interstate highway expansion programs. Focusing on the pre-highway expansion period, our paper estimates the effects of internal trade barriers on regional gains from external integration.

The second strand of literature investigates the role of internal geography in explaining the uneven regional effects of international trade. Coşar and Fajgelbaum (2016) introduce internal geography into an international trade model to show that regions closer to ports in China specialize in export-oriented sectors, while interior regions do not engage in international trade. Xu and Yang (2021) show that improving international port access enhances welfare gains for interior regions in China. Fajgelbaum and Redding (2022) provide a theoretical framework to disentangle the effects of internal and external trade cost reductions on the spatial distribution of economic activity. We contribute to this body of evidence by analyzing the spatial effects of a large-scale trade liberalization episode.

The third strand examines the effects of trade on structural change. Belmar (2023) documents positive effects of trade on structural transformation, using evidence from Colombia where certain regions gained access to international markets following the opening of the Panama Canal. In contrast, Cravino and Sotelo (2019) find negative effects of reduced trade costs on manufacturing employment. Our paper contributes further evidence on the positive impacts of trade on structural change. In the context of India, Fan et al. (2023) highlight that structural change has largely bypassed the manufacturing sector, driven by productivity growth in the service sector. We show that, while manufacturing employment may not appear to have grown significantly at the aggregate level, these aggregate trends obscure substantial regional heterogeneity that existed until 2005.

Lastly, this paper contributes to a vast body of literature on India's trade liberalization. Topalova and Khandelwal (2011) show that reductions in input and output tariffs significantly boosted firm-level productivity. Goldberg et al. (2010a) show that access to new imported inputs enabled firms to expand their product scope, accounting for a quarter of the growth in manufacturing output between 1991 and 1997. Topalova (2007) examines the local labor market effects of trade liberalization on poverty, comparing districts with varying levels of tariff exposure. We extend this literature by investigating spatial heterogeneity across districts based on their international access and providing new evidence on the regional impacts of trade liberalization.

The remainder of this paper is organized as follows: Section 2 provides background about India's trade liberalization and internal geography and discusses the mechanisms at play. Section 3 describes all the datasets used in this paper. Section 4 outlines the empirical strategy and discusses the results. Section 5 concludes.

2 Background

2.1 India's Trade Liberalization

Prior to 1991, India operated like a closed economy pursuing import substitution-led industrialization. Its policy was characterized by heavy industrial licensing and strict import controls in the form of high tariff and non-tariff barriers. In 1987, the average tariffs stood at 95%, and 87% of the products were subject to quantitative restrictions. An actual user policy on imports further restricted the imports of intermediate products (Topalova, 2007). Towards the end of the 1980s, a combination of external and internal crises paved the way for India's trade liberalization. The Gulf War in 1990 led to a steep increase in oil prices and a reduction in remittances from Indian workers abroad. Political instability led to significant capital outflows, resulting in a severe balance of payments crisis. By June 1991, India's foreign exchange reserves were sufficient to cover only two weeks of imports.

To avoid defaulting on its balance of payments, India sought a loan from the International Monetary Fund (IMF) which was granted conditional on India undertaking some macroeconomic reforms. These reforms were completely unanticipated and externally enforced. Trade liberalization formed a huge part of these reforms. Import licensing on virtually all intermediate inputs and capital goods was abolished. Within the first five years of the reforms, average tariffs fell by about 50 percentage points (pp). The actual user policy on imports was also eliminated. As a result, overall imports rose by 130% between 1987 and 2000. This growth was mostly driven by imports of intermediate inputs which grew by 227% during the same period (Goldberg et al., 2010b). Overall, trade liberalization resulted in significant gains for the manufacturing sector, enhancing firm-level productivity and expanding product variety (Topalova and Khandelwal, 2011; Goldberg et al., 2010a). Figure 1 shows the decline in tariffs and effective rates of protection[3] in India form 1987-2008.

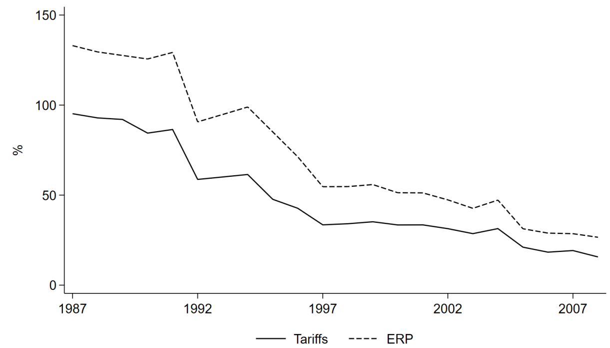
In this paper, we add to the evidence on trade liberalization by introducing spatial heterogeneity across regions in India in terms of their access to international gates and examine how firm and labor market outcomes respond to trade liberalization in the presence of internal trade barriers.

2.2 Internal Geography of India

The bulk of India's international trade passes through twelve major seaports and two airports. For instance, in 1998-99, seaborne trade accounted for 65% of India's total trade by value and twelve major seaports handled 87.5% of the total cargo at seaports (Basic Ports Statistics, 2001-02). The two major airports in Delhi and Mumbai handled 71% of the total airborne freight. Another important fact that highlights the importance of international gates in India is the low volume of trading with its neighbors. Trade with countries sharing a border does not necessarily require cargo to pass through ports as it can ship directly across countries via roads. For example, in the year 2000, India's total imports

[3]See section 4.1 for a discussion on Effective Rate of Protection.

Figure 1: Trade Liberalization in India



Note: The figure shows the average decline in output tariffs and effective rates of protection from 1987-2008

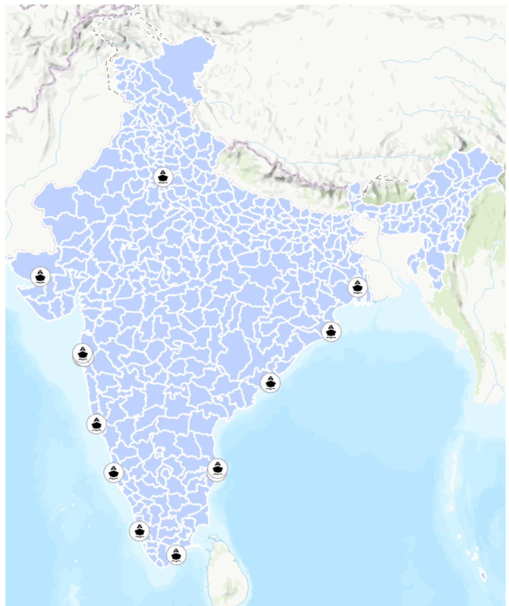
from neighbors accessible via roads including Pakistan, Bangladesh, Afghanistan, Nepal, Myanmar, and Bhutan amounted to only 1.21% of total imports.

Owing to India's diverse geography, not all regions have equal access to international gates[4]. As shown in Figure 2, aside from the international airports in Delhi, most interior and non-coastal states lack direct access to international gates, creating significant barriers for these regions to engage in international trade. In the 2000s, the Government of India made substantial infrastructure investments to address these challenges, including the construction of the 5,800 km Golden Quadrilateral highway, aimed at improving connectivity to major ports. Despite these efforts, internal trade barriers remain significant. According to Van Leemput (2021), internal barriers accounted for 51% of total trade barriers in non-port states in India during 2011-12. In the 1990s, which also happens to be the period of India's trade liberalization, the state of India's road networks was worse. In the 1990s, the time to reach a port from a district at the 90th percentile was 8 times larger than a district at the 10th percentile.

In the presence of high internal travel costs, internal geography can become a source of regional inequalities in an open economy. In Figure 3, we plot district-level output growth in the manufacturing sector against the travel time to ports for each district. The travel time is computed from a district's centroid to the nearest international gate as per the 1996 road network. As shown in Figure 3(a), before 1992, when India was largely a closed economy, output growth shows no correlation with international access. However, following India's trade liberalization, there is a significant negative correlation between travel time and output growth as shown in Figure 3(b). From 1992-04, access to international markets becomes a lot more important for manufacturing output.

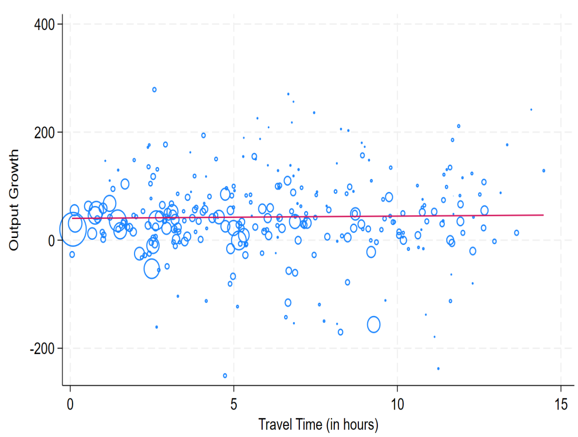
[4]We use the terms international gates, gates, and ports interchangeably.

Figure 2: Major Sea and Air Ports in India

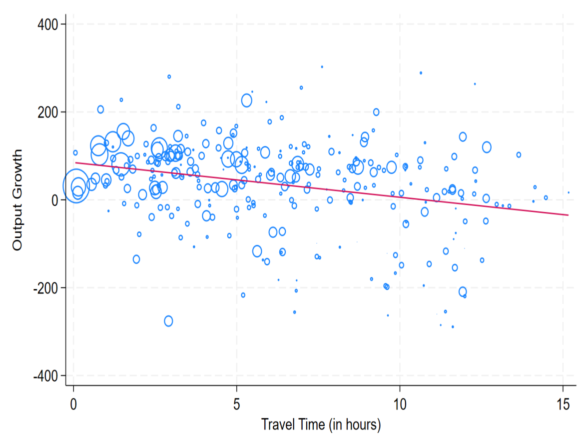


Note: Figure depicts the 12 major seaports and two major airports in India. Shapefile depicts districts as per 1991 boundaries.

Figure 3: Manufacturing Output Growth and Proximity to Ports



(a) 1987-1992



(b) 1992-2004

Note: The figure shows the correlation between international access and output growth in pre and post-liberalization period. Top and bottom 2.5% sample is trimmed in both periods to remove outliers. The fitted line represents the linear fit from the regression using log of real output in the base period as weights.

2.3 Mechanisms

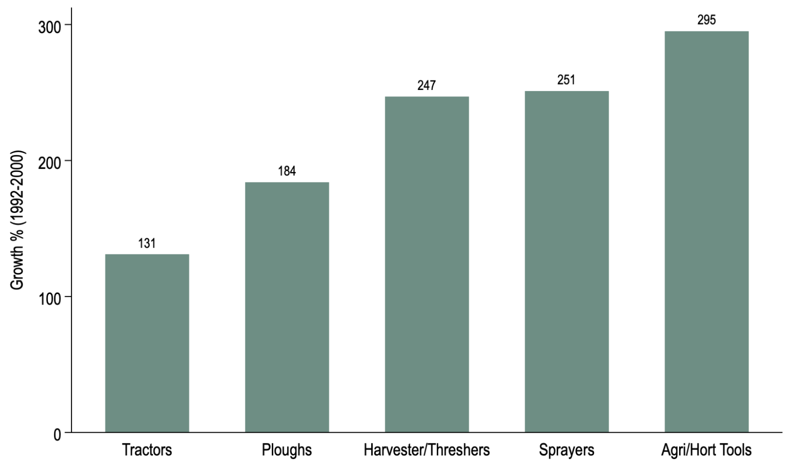
Trade liberalization can affect firms via several channels. A reduction in tariffs on intermediate products can improve access to better and cheaper inputs, boosting productivity. Increased foreign competition due to a reduction in output tariffs can act as a disciplining device leading to improved firm efficiency. Lastly, exposure to technologically superior goods in both the input and output market can lead to technology spillovers on local firms. However, increased foreign competition can also adversely affect some firms, potentially forcing some to exit the market (Melitz, 2003). Goldberg et al. (2010b) show that in India, tariff reductions led to a 227% increase in intermediate imports between 1987 and 2000, leading to significant gains in firm productivity and an expansion in product scope (Topalova and Khandelwal, 2011). But in the presence of high internal trade costs, the location of firms becomes an important factor for the transmission of these effects.

Most imports enter a country via an international gate and then get shipped to domestic markets. As internal shipping is costly, for a given imported input, a firm located near a port can access the input at a cheaper price as compared to an inland firm. Simultaneously, for a given final good a firm located near a port faces greater competition from imported final goods as compared to an inland firm as foreign prices lose competitiveness with high internal trade costs. The spatial effects of trade liberalization extend to the export market as well. Higher productivity gains from cheaper imported inputs and lower shipping costs to international markets might give a price advantage to firms near ports over inland firms in the export market. Thus, overall trade liberalization can lead to ambiguous spatial effects on manufacturing activity.

Given these internal trade costs, trade liberalization can have spatial effects on structural transformation within an economy. If districts with better international access experience larger gains in manufacturing activity due to trade liberalization, they would experience a higher labor demand in the manufacturing sector relative to districts with poor international access. Furthermore, since wages in manufacturing typically exceed those in agriculture, these districts may see larger income gains. Under the assumption of non-homothetic preferences, income gains in turn lead to a greater demand for services in these districts.

Adoption of labor-saving technology in agriculture can increase agriculture productivity and potentially release labor from agriculture (Cheung and Yang, 2024). Figure 4 shows that imports of agriculture equipment like tractors, ploughs, sprayers, etc., saw a massive increase between 1992-2000. Internal trade barriers can influence structural change by affecting the rate of adoption of these labor-saving technologies as districts that are more integrated with international markets are better positioned to import agricultural equipment. Thus, understanding the importance of international access for the transmission of gains from trade liberalization is critical to unpacking the effects of trade liberalization on structural change in the economy.

Figure 4: Growth in Imports of Agriculture Inputs (1992-2000)



Note: Figure depicts the growth in value of imports for tractors, ploughs, harvesters/threshers, sprayers, and other agriculture/horticulture tools between 1992 and 2000. Data taken from UN COMTRADE.

To quantify these spatial effects of trade liberalization on manufacturing activity and structural transformation, we focus on the period from 1987 to 2004. Although we have data until 2012, we restrict our long-run analysis to 2004 as the period from 2005-12 coincides with the passage of two large-scale national policies in India which could potentially be correlated with a district's international access and its economic outcomes. First, India passed the Special Economic Zones (SEZ) Act in 2005 which promoted the establishment of hubs specializing in the production of export-oriented goods. Between 2005-13, 147 SEZs became operational out of which more than 90% were located in the coastal districts. Studies have found positive effects of SEZs for regional structural change (Gallé et al., 2022; Hyun et al., 2018). Second, India enacted a massive rural employment guarantee program, MGNREGA in 2006 which offered 100 days of employment per eligible household in a financial year. It largely offered public works jobs like road building and water management. Studies show that MGNREGA implementation has been heterogeneous across states and most employment generation until 2009 was concentrated in inland states of Rajasthan, Madhya Pradesh, and Chhattisgarh (Dreze and Oldiges, 2009; Sukhtankar, 2017). Finally, while Phase I of the previously discussed major highway expansion program in India was almost complete by 2006, Phase II was launched in 2006. These constructions can drastically change the international access for far-off districts. By restricting the data to earlier periods, we ensure that these large-scale highway projects do not create a measurement error in travel times.

3 Data

International Access: The key variable in this study is the access of districts to international markets via major international sea and airports. Our measure of international access is computed as an iceberg trade cost[5]. To measure trade costs, we need a measure of travel time from district centers to ports. To accurately measure travel time, we digitize the 1996 detailed road map of India published by the Government of India [6] with the assistance of a digital mapping company, CyberSwift. This allows us to identify five types of roads - National and State Highways, All-Weather Motorable Roads, Fair-Weather Motorable Roads, Roads Where Delays May Occur, and Other Roads. To assign average speeds to each of these road types, we rely on the values used by Allen and Atkin (2022), Vogel et al. (2024) and World Bank Report (2002). The report estimates an average speed of 35-40 kmph on highways in India. Allen and Atkin (2022) and Vogel et al. (2024) assign speeds of 40 mph for highways and 20 mph for the lowest quality roads. To be consistent, we assign speeds of 30 mph for highways, 20 mph for all-weather motorable roads, and 10 mph for all other road types[7]. Panel A and B of Figure 5 show the placement of highways and all-weather motorable roads in 1996. We get the district boundaries of India using the 1991 shapefiles from IPUMS International. For international gates, we obtain geocodes of the 12 major seaports and 4 major airports in India from the web. To measure international access of a district, we start by computing the travel time from the district's centroid to its nearest international gate.

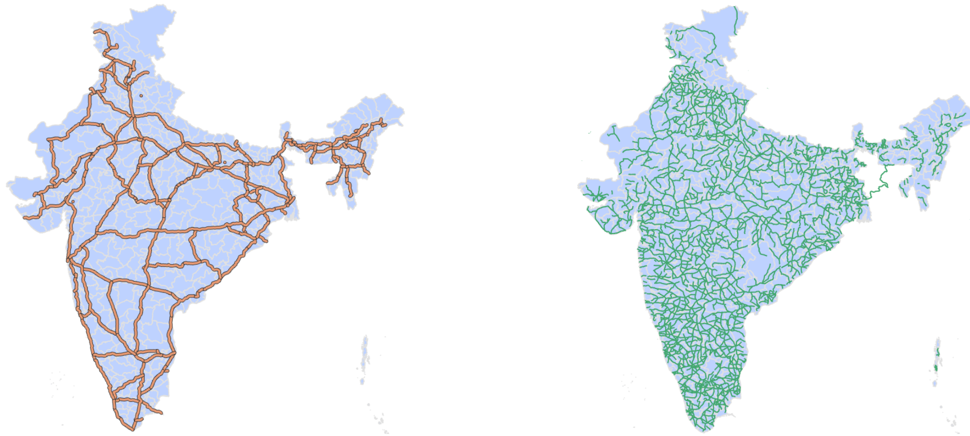
Given the assumed speeds, we apply Dijkstra's algorithm to find the shortest path between the district's centroid and its nearest gateway. The algorithm breaks the road network into a set of nodes and edges, where the edges represent travel time on the road segment between the nodes. By using the travel times as weights, and summing over different possible routes, the algorithm finds the shortest path between origin and destination points. Since both a district center and ports may not be directly connected to a road network, we assume that a connection exists between these points to the nearest road segment. The results from the exercise are summarized in Table 1. For the 340 districts used in our analysis, the average and median travel times are 9.01 hours and 8.75

[5]Intuitively, iceberg trade cost are > 1 and capture the units of a good that need to be shipped from a district center to a port to deliver 1 unit of the good at the port.

[6]We thank Treb Allen for sharing the older raw road maps of India which are no longer publicly available.

[7]Results of this paper are robust to changing the average speed on each road by ± 5 mph.

Figure 5: Highways and All-Weather Motorable Roads



(a) Highways

(b) All-Weather Motorable

hours, respectively. The interquartile range is 7.73 hours, and the farthest district is 18.19 hours away from a port. Figure 6 shows the heat map of the districts based on their calculated travel time from international gates. Both Table 1 and Figure 6 highlight that there is considerable variation across districts in their travel time to international gates. Figure 7 further highlights the importance of using the full road network to measure travel costs as opposed to the geographic distance between two points. Both Bharuch and Aurangabad (district centroids) are roughly 170-175 miles away from Mumbai. But Bharuch being closer to the highways takes 2.6 hours less to travel to Mumbai as compared to Aurangabad. A reliance on geographic distances does not take into account a district's connectivity via different types of roads and, thus, does not capture the proximity between district pairs or between districts and international gates accurately.

Using the travel times, we compute the iceberg trade cost as follows:

$$Trade\,Cost_{od} = 1 + \gamma(Travel\,Time_{od})^{\rho}$$

where o is the origin (district centroid) and d is the destination (nearest port). Following Barnwal et al. (2024), Baum-Snow et al. (2020), and Alder (2016), we set $\rho = 0.8$ and calibrate γ such that the median trade cost between any two *district centers* is 1.25 on the 1996 road network[8]. We summarize trade costs in Table 1. We obtain a median trade cost of 1.10 which implies that a travel time of 8.75 hours to a port results in a loss of 10% of the value shipped. The closest district to a port experiences no loss while the farthest

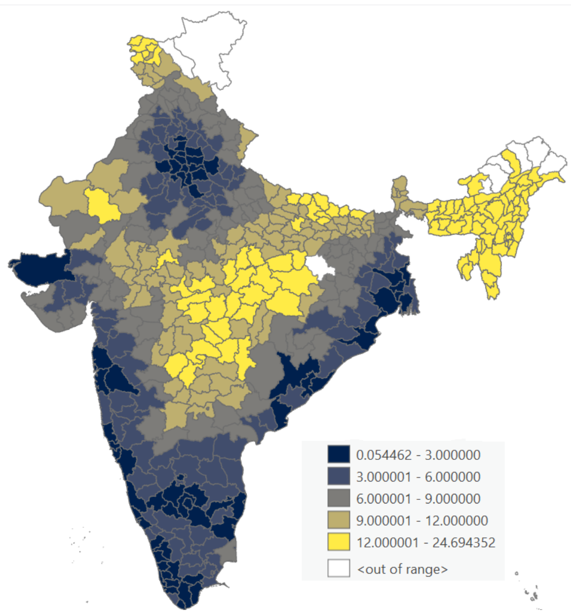
[8]The results in this paper are robust to setting $\gamma = 0.004$ as used by Baum-Snow et al. (2020). Results are also robust to a range of γs calibrated such that the median trade cost ranges between [1.20-1.30].

district loses 20%.

Table 1: Descriptive Statistics - International Access

	Mean	S.D.	25th pctile	Median	75th pctile	Min	Max
Travel Time (in hours)	9.25	4.81	5.39	8.88	12.79	0.11	21.05
Trade Cost	1.10	0.04	1.07	1.10	1.13	1.00	1.20
# Districts	340						

Figure 6: Heat Map of Travel Time across Districts

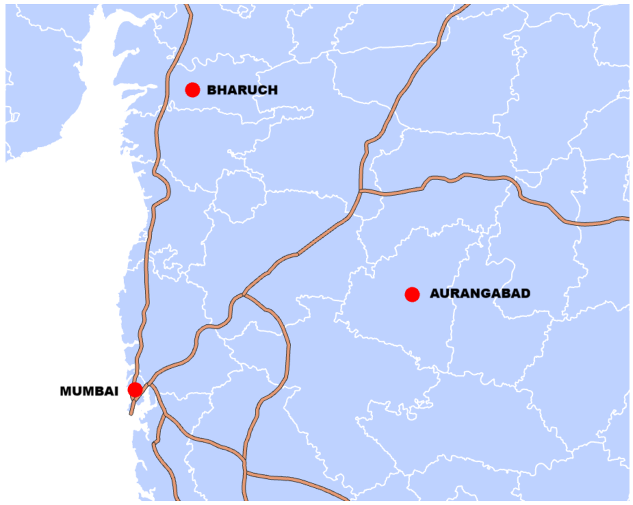


Note: The figure uses all the districts in India in 1991. Travel time is divided into 5 bands. In our analysis, we work with a sample of 340 districts. Districts from North-East states and Jammu & Kashmir are dropped

We do not focus on railways as roads play a much larger role in the movement of freight in India. In 2000-01, roads accounted for 61.3% of the total freight movement (GoI). The share in non-bulk commodities is much larger. Further, as we are interested in travel times to ports, rails in India do not offer any advantage over road network for smaller distances, usually up to 500km (Sahu et al., 2022). Approximately, 95% of our sample districts are under 600 km from ports. In our sample, we further drop North-East states which are farthest from ports and more likely to rely on rails for freight.

Annual Survey of Industries: Our primary source of data on manufacturing activity comes from the Annual Survey of Industries (ASI) conducted by the Ministry of Statistics

Figure 7: Importance of Road Network in Measuring Travel Time



Note: Both Bharuch and Aurangabad are approximately 170-175 miles away from Mumbai. The travel time from Bharuch to Mumbai using the road network is 7.07 hours while that of Aurangabad to Mumbai is 9.67 hours.

and Programme Implementation (MoSPI), Government of India. The ASI is a representative survey of Indian manufacturing establishments in the organized sector[9]. It covers all large establishments (≥ 100 workers) and smaller establishments with, roughly, one-third probability. We use annual data on sales and labor employed by establishments for this analysis.

In terms of geographic identifiers, ASI provides a district code but withholds information on district names. As district boundaries change over time, district names are crucial to consistently identify a district over time. We augment the ASI data from 1987-88 to 2009-10 with consistent district identifiers using all available publications on Indian districts released by MoSPI. Details of the exercise are provided in Appendix A.1. Our final dataset consists of 345 districts at the 1987 district boundaries. We aggregate the ASI data to create an $industry \times district$ panel for the manufacturing sector at the 4-digit National Industry Classification (NIC) level of 1998.

Economic Census: A majority of India's labor force is employed in the unorganized sector. To get a complete picture of the manufacturing sector's response to employment, we use data from the Economic Census. The Economic census covers all establishments in the organized as well as the unorganized sector. We focus on the rounds conducted in 1990, 1998, and 2005. We use the Economic Census to analyze the total number of firms

[9]The organized sector comprises all firms that are registered with the government under the Factories Act of 1948. These are establishments that employ more than 10 workers and use electricity or 20 workers and do not use electricity

and labor employed at the $industry \times district$ level.

National Sample Survey: Our data on labor market outcomes comes from the National Sample Survey (NSS) and Population Census of 1991. The NSS are nationally representative household surveys conducted every five years. They provide data on employment status, industry of employment, and demographics at the individual-level. Our key outcome variable, sectoral employment share is defined as the total employment in a sector as a share of the district's labor force. We use NSS rounds from 1987-88, 1999-00, and 2004-05[10] and aggregate each round to obtain the district-level share of labor force for agriculture, manufacturing and services using the sample weights provided by the NSS. For the pre-liberalization period, we supplement this data with Population Census data from 1991. The Population Census provides district-level data on employment disaggregated by industry of employment.

Tariffs: We obtain HS6 product-level tariff data from Topalova (2007) for the period 1987 to 2001. For the period 2002-05, we obtain tariff data from the UNCTAD-TRAINS database hosted on the World Integrated Trade Solutions (WITS) portal. This combined dataset contains tariffs for 5,045 HS6 products. Using product concordances from WITS, we first map the HS6 product codes to ISIC Revision 3 which maps one-to-one with India's NIC 1998 industry codes. We collapse the tariffs at the 4-digit NIC 1998 industry-level to generate average industry tariffs. Next, we combine these industry tariffs with India's Input-Output Matrix of 1993-94 to derive input tariffs[11]. Our final dataset consists of tariffs from 125 4-digit NIC 1998 manufacturing industries.

Migration: The primary source of migration data is the Census of India for 2001. An individual is recorded as a migrant if the place in which she is enumerated during the Census is different than her place of immediate last residence as recorded in the previous Census in 1991. Publicly available data only contains the destination district and whether the migrants' origin is within the same state or outside. We obtain district-to-district migration flows from Ghose (2021) that contain both the destination and the origin district of the migrant[12].

[10]We exclude NSS 1993-94 from our analysis as it lacks district-level identifiers.

[11]Input tariffs, τ^{ip} are computed as follows: $\tau^{ip}_{jt} = \sum_i \alpha_{ij}\tau_{it}$. Weights α_i are the share of input i in the total value of industry j's output and τ is the tariff on input i in year t. Weights α are taken from India's Input-Output Table (1993-94) after mapping inputs to 4-digit NIC 1998 industry codes.

[12]We are grateful to Devaki Ghose and Kerem Coşar for generously sharing the data. This data was originally obtained through a special agreement with the Government of India.

4 Empirical Strategy and Results

4.1 Manufacturing Activity: Intensive Margin

The empirical strategy in this section exploits the differences in districts' access to international markets to identify the effects of trade liberalization on an industry's output across districts. The positive effects of lower input tariffs can be offset by increased competition from lower output tariffs or supplemented with the positive effects from the disciplining channel due to lower output tariffs. In order to measure the net effect of trade liberalization on protection in an industry, we create the *effective rate of protection (erp)*, as defined by Corden (1966),

$$erp_{jt} = \frac{Output\,tariff_{jt} - Input\,tariff_{jt}}{1 - \sum_s \alpha_{js}}$$

where α_{js} is the share of input from industry s in the value of output of industry j. This measure has been widely used to measure protection in industries [13]. Input tariff follows,

$$Input\,Tariff_{jt} = \sum_s \alpha_{js} \cdot Output\,Tariff_{st}$$

In some specifications, we will also use both input and output tariffs separately to better understand the channels driving the results.

We start by examining total manufacturing output and labor productivity. We use sampling weights to aggregate establishment-level data from ASI at the district-industry-year level. We only keep 4-digit manufacturing industries in our sample. The baseline specification takes the form:

$$y_{jdt} = \beta_0 + \beta_1\, erp_{jt} \times Trade\,Cost_d + \beta_2\, erp_{jt} + DMA_{dt} + X_{jd,pre}\,\delta_t + \lambda_{jd} + \delta_t + \epsilon_{jdt} \qquad (1)$$

Here, y is the outcome variable for industry j in district d in year t. We study two outcome variables: (1) log of sales (output) and, (2) log of labor productivity of industry j in district d, defined as the output per worker. Our sample covers the period from 1991-92 to 2004-05. We deflate sales by using industry-level Wholesale Price Index for 1993-94. erp is the effective rate of protection faced by industry j in year t. $Trade\,Cost$ measures the iceberg trade cost of moving goods between the district centroid and the nearest international gate, using the 1996 road network. Year fixed effects, δ_t control for aggregate shocks and *industry-district* fixed effects, λ_{jd} control for time-invariant industry-

[13] Specifically in the Indian context in Topalova and Khandelwal (2011)

district unobservables. DMA_{dt} is the domestic market access of the districts following Donaldson and Hornbeck (2016) to control for the possibility that districts closer to the ports have better access to domestic markets that boosts output[14]. $X_{jd,pre}$ contains industry j's pre-liberalization level of y in district d interacted with the year fixed effect and pre-liberalization share of an industry's output in a district's total output interacted with year fixed effects. This allows the district-industry pairs to follow a different growth path based on the initial size and importance of the industry for the district. We define the pre-liberalization period as 1987-91. Standard errors are two-way clustered at the 4-digit industry-year and district level.

The coefficient of interest, β_1, captures the importance of international access in comparing output growth across districts resulting from falling effective protection within an industry. $\beta_1 > 0$ implies that for a given decline in *erp*, the output within an industry grows relatively more in districts with a lower trade cost or better international access.

β_1 is identified by comparing two districts within a 4-digit industry, with the same pre-liberalization levels of outcome and facing the same *erp* change, but one of them is more integrated with international markets than the other. Since our comparison is within an industry, the endogeneity of tariff cuts is not a concern for us. Irrespective, a large literature studying India's trade liberalization provides enough evidence that liberalization was exogenous across industries, at least till the late 90s (Hasan et al. (2007); Varshney (1998); Topalova and Khandelwal (2011); Goyal (1996)). A possible threat to identification would be if any major ports were built post-1991, perhaps, in regions better suited for future growth. This is not the case as most of the major ports were built during the British rule to facilitate trade. The only exception is the Kamarajar port in Chennai (2002), but since Chennai International Airport (established in the 1960s) is already included as an international gate, excluding Kamarajar port from the analysis keeps the results unchanged. As described below, the results are further robust to restricting the sample to the year 2000.

The remaining potential threats arise from other changes during the sample period that can simultaneously affect manufacturing outcomes and are correlated with the proximity of a district to international gates. Tariff cuts were just one component of a broad liberalization package. The second major reform was the automatic approval of Foreign Direct Investment (FDI) up to 51% of a firm's equity in several manufacturing industries. If FDI flows are larger in districts with better international access, this can bias our results. Using data on FDI inflows from CMIE's CapEx database, we find that FDI flows are indeed highly spatially concentrated, with 55% of total FDI projects in the period 1993-2008

[14]See A.2 for details on the construction of *DMA*.

going to 11 districts. Unsurprisingly, these are amongst the most urban districts in the country at the time[15]. Some of these districts like Mumbai, Delhi, Bangalore, and Kolkata are also port districts. We show that our results are unchanged when these high FDI receiving districts are eliminated from the sample. This further addresses the concern that the benefits of liberalization might accrue more to urban districts that could also be located closer to the international gates.

Finally, a possible threat is India's ambitious highway expansion program that was ongoing throughout the 2000s. The first and most important is the Golden Quadrilateral (GQ), whose construction began in 2001. GQ is a network of high-speed roads connecting the four major cities - Delhi, Mumbai, Chennai, and Kolkata. Several papers have documented the benefits of GQ for manufacturing activity in districts along the GQ network (Asturias et al., 2019; Ghani et al., 2016). A large part of the GQ passes through regions that are invariably closer to international gates, as shown in Figure 8. Although GQ came close to completion in 2006 and our sample is limited to 2004, it can bias still the results for some later years if these districts gain due to GQ and better access to inland domestic markets. To control for the potential bias from GQ, we show that restricting our analysis to the pre-2001 period leaves our results unchanged. Finally, we conduct placebo tests using the period 1987-1991 to show that our results are not driven by any pre-existing spatial trends in the outcomes.

Results: In Table 2, we summarize the results for manufacturing output. The results from Equation 1 are presented in Column (1). The estimate of β_1 is positive and significant. This shows for a given decline in the effective rate of protection in an industry, districts with better international access witnessed higher output growth. In particular, the coefficient implies that for a 10 pp decline in *erp*, reducing the trade cost by the median value of 0.1, leads to 2.56% $(= 2.564 \times (1 - 1.1) \times -10)$ larger output. This confirms that international access is important in the regional distribution of gains from trade liberalization. In Columns (2)-(6), we add a host of controls to the baseline specification to check the robustness of our estimate.

In Column (2), we restrict our sample to 2000, the pre-GQ period. In Column (3), we exclude the 11 districts with 55% FDI share from our sample. In Column (4), we control for state-year fixed effects to account for state-level policy variations. This controls for state-specific policies that might differ over time on important matters such as labor laws, that could potentially affect manufacturing outcomes. In Column (5), we add industry-year fixed effects to account for any industry-level shocks, such as de-licensing

[15]These districts are Delhi, Mumbai, Bangalore, Kolkata, Chennai, Hyderabad, Pune, Gurgaon, Ghaziabad, Chengalpattu, and Thane

Figure 8: Golden Quadrilateral

Note: Figure shows the placement of GQ. Large parts of GQ pass through coasts and connect to Delhi.

of an industry. This fully absorbs the coefficient on *erp* but our coefficient of interest, β_1, is still estimated. Finally, to address the possibility of the results being driven by pre-existing trends in industry-district output, we control for the average output growth of an industry-district pair in the pre-liberalization period interacted with year-fixed effects in Column (6). The results in Table 2 indicate that our estimates of β_1 stay positive, statistically significant and stable across specifications. Economically, for a 10 pp decline in the effective rate of protection, the causal effect on output of reducing trade cost by the median ranges from 2.19% to 3.03%. From 1990 to 2005, *erp* declined on average by 98 pp. Using our baseline estimate, this implies that better international access at the median, led to on average a 25.1% higher manufacturing output in an industry with the average erp decline relative to an industry with no change in protection.

As a separate exercise, we conduct a falsification test by estimating Column (4) of Table 2, with all the baseline controls and state-year fixed effects, for the pre-liberalization period, i.e., from 1987 to 1991. The results are reported in Column (1) of Table 4 and show that no significant pre-liberalization trends are driving our main results. We also allow our main results to be semi-parametrically estimated in international access. We split the trade costs into quintiles. We find that relative to the 5th quintile (farthest from ports), all the quintiles gain more, with gains monotonically increasing in reduced trade costs (Figure 9).

To understand if the output or input tariff reductions matter more for the spatial effects of trade liberalization as seen above, we run our baseline specification with input and output tariffs separately in Table 5. Column (1) shows that for the period of 1990-2005, districts with lower trade costs gained more primarily due to declining output tariffs. This implies that pro-competition effects are stronger than the intermediate input channel. This is slightly in contrast to the previous literature documenting stronger aggregate positive effects from the import channel (Topalova and Khandelwal (2011); Goldberg et al. (2010a)). To dig deeper, we restrict our sample to 1998 to closely resemble the time period used in these studies. The results from the curtailed sample align with the literature. Column (2) shows that, indeed, the input channel is much stronger and significant during the first phase of liberalization, even spatially. Thus, our evidence points out that during the 15 years of our sample, both input and output tariffs matter for diverse regional gains, with input tariffs being more important in the early years of liberalization and output tariffs becoming more important later.

Lastly, to address any concerns regarding the effects coming from specific years, we show the dynamic effects of trade liberalization in Figure 10. We take 1992 as the year of the liberalization. A small caveat with this choice is that while most industries saw major tariff declines in 1992, some saw delayed cuts beginning mid-90s. With that in mind, the results confirm that districts with lower trade costs experience additional manufacturing gains from 1993 onward and the gains are persistent for 10 years afterwards.

Next, we summarize our results for labor productivity in Table 3. Column (1) is the baseline specification where we further control of industry's pre-liberalization level of labor productivity in the district interacted with year fixed effects. The point estimates show that for a 10 pp decline in an *erp*, median reduction in trade cost between a district centroid and the nearest port leads to 1.48% increase in labor productivity. In Columns (2) to (6), we repeat the robustness checks undertaken in Table 2 for output and find that our effect size is stable and significant across specifications. Given the average *erp* decline of 98 pp from 1990-2004, better international access at the median, leads to on average 14.5% higher labor productivity. Thus, we find substantial evidence that the benefits of trade liberalization for a country's manufacturing sector are unequally distributed in space, with international access playing a crucial role.

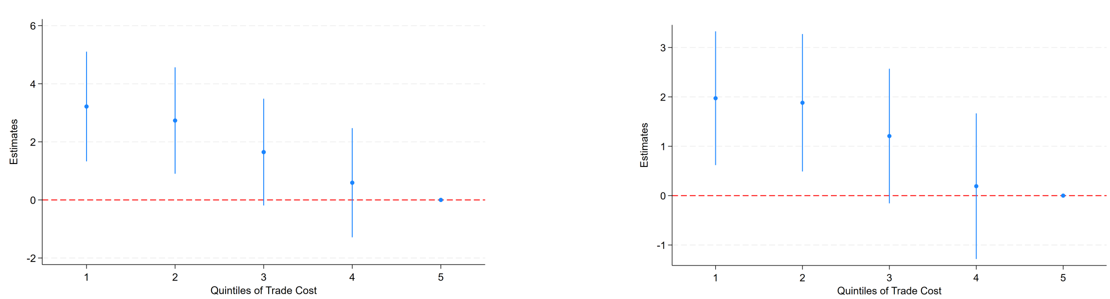
Column (2) of Table 4 shows the results from the falsification test for pre-trends in labor productivity. Again, we find no evidence of any significant pre-liberalization trends. Semi-parametric estimation in Figure 9, further, shows that the gains from reduced protectionism are monotonically increasing with decreasing trade cost bands.

Table 2: International Access and Manufacturing Output

	Log Output					
	(1)	(2)	(3)	(4)	(5)	(6)
erp	-2.564***	-2.192***	-3.034***	-2.454***	-	-2.407***
	(0.721)	(0.673)	(0.705)	(0.677)		(0.855)
$erp \times Trade\ Cost$	2.443***	2.094***	2.878***	2.345***	2.334***	2.295***
	(0.670)	(0.626)	(0.654)	(0.627)	(0.612)	(0.794)
Observations	70,517	52,181	62,860	70,517	70,505	54,546
R-squared	0.776	0.802	0.772	0.779	0.785	0.774
District-Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	N	N	Y
State-Year FE	N	N	N	Y	N	N
Industry-Year FE	N	N	N	N	Y	N

Note: The dependent variable is the log of real sales of industry j in district d. erp is the effective rate of protection faced by the industry in year t. *Trade Cost* measures the iceberg trade cost between a district's centroid and its nearest port. Column (2) restricts the period of analysis up to the year 2000 to exclude the effects of the GQ highway. Column (3) drops the 11 districts that received most of the FDI during 1993-2008. Column (4) includes state-year fixed effects. Column (5) includes industry-year fixed effects. Column (6) controls for the average growth in output for industry j in district d during 1987-91 interacted with year fixed effects. Standard errors are two-way clustered at the 4-digit industry-year and district level. The sample consists of 101 4-digit industries across 337 districts. ***, **, * show significance at 1%, 5%, and 10%, respectively.

Figure 9: Semi-Parametric Estimates of Trade Cost



(a) Real Output **(b)** Labor Productivity

Note: This figure reports the results from a semi-parametric estimation of Equation 1. *Trade Cost* is divided into quintiles. Each point estimate represents the additional effect on output/labor productivity for the i'th quintile relative to the 5th quintile for a 10 pp decline in erp.

Table 3: International Access and Manufacturing Labor Productivity

	Log Labor Productivity					
	(1)	(2)	(3)	(4)	(5)	(6)
erp	-1.482***	-0.952**	-1.537***	-1.523***	-	-1.467***
	(0.402)	(0.386)	(0.446)	(0.435)		(0.464)
$erp \times Trade\ Cost$	1.384***	0.889**	1.440***	1.425***	1.473***	1.357***
	(0.376)	(0.360)	(0.414)	(0.407)	(0.358)	(0.434)
Observations	64,785	46,433	57,716	64,785	64,774	50,106
R-squared	0.638	0.684	0.642	0.641	0.657	0.635
District-Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	N	N	Y
State-Year FE	N	N	N	Y	N	N
Industry-Year FE	N	N	N	N	Y	N

Note: The dependent variable is the log of labor productivity of industry j in district d. Labor productivity is obtained by dividing real sales of industry j in district d by the total labor employed by j in d. erp is the effective rate of protection faced by the industry in year t. *Trade Cost* measures the iceberg trade cost between a district's centroid and its nearest port. Column (2) restricts the period of analysis up to the year 2000 to exclude the effects of the GQ highway. Column (3) drops the 11 districts that received most of the FDI during 1993-2008. Column (4) includes state-year fixed effects. Column (5) includes industry-year fixed effects. Column (6) controls for the average growth in output for industry j in district d during 1987-91 interacted with year fixed effects. Standard errors are two-way clustered at the 4-digit industry-year and district level. The sample consists of 101 4-digit industries across 337 districts. ***, **, * show significance at 1%, 5%, and 10%, respectively.

Table 4: Placebo Test: International Access and Manufacturing Activity

	Output	Labor Productivity
	(1)	(2)
erp	0.300	0.806
	(1.019)	(0.739)
$erp \times Trade\ Cost$	-0.421	-0.812
	(0.933)	(0.677)
Observations	23,284	23,250
R-squared	0.778	0.722
District-Industry FE	Y	Y
State-Year FE	Y	Y

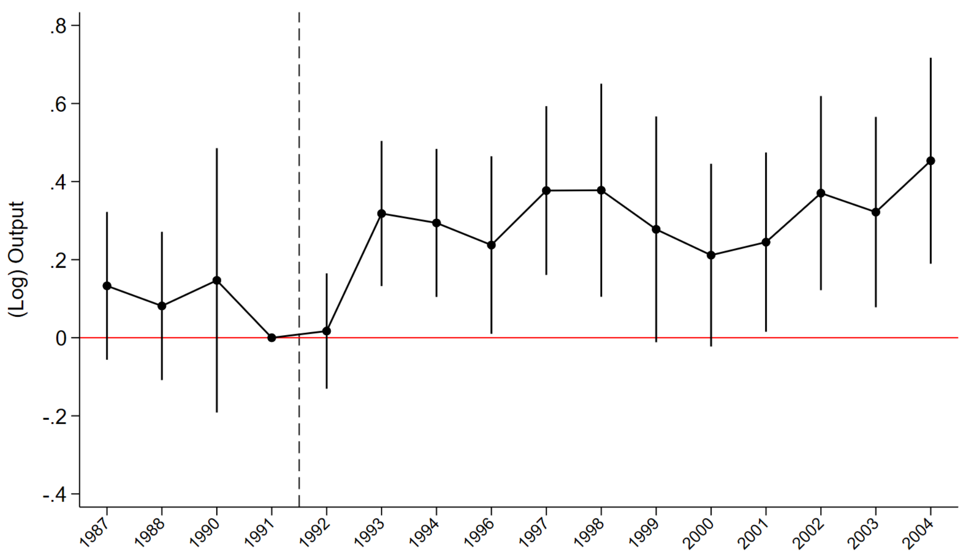
Note: The dependent variable in Column (1) is real sales and in Column (2) is labor productivity. Labor productivity is obtained by dividing real sales of industry j in district d by the total labor employed by j in d. The dependent variables are in logs. erp is the effective rate of protection faced by industry j in year t. *Trade Cost* measures the iceberg trade cost between a district's centroid and its nearest port. Standard errors are two-way clustered at the 4-digit industry-year and district level. The sample covers the pre-liberalization period of 1987-91 and consists of 101 4-digit industries across 337 districts. ***, **, * show significance at 1%, 5%, and 10%, respectively.

Table 5: Disentangling the Effects of Input and Output Tariffs - Output

	Log Output	
	(1)	(2)
$Input\ Tariff \times Trade\ Cost$	0.485	12.329***
	(4.537)	(3.841)
$Output\ Tariff \times Trade\ Cost$	4.941***	0.750
	(1.795)	(1.420)
Observations	69,680	40,245
R-squared	0.776	0.838
District-Industry FE	Y	Y
Year FE	Y	Y

Note: The dependent variable is log of real sales. Column (1) contains results from the entire sample running from 1991 to 2004 while Column (2) restricts the sample to 1991-97. *Output Tariff* is computed by aggregating HS6 product-level tariffs to 4-digit NIC 1998 industry-level. *Input Tariff* is computed using the Input-Output Table of 1993-94 published by the Government of India. *Trade Cost* measures the iceberg trade cost between a district's centroid and its nearest port. Standard errors are two-way clustered at the 4-digit industry-year and district level. The sample consists of 101 4-digit industries across 337 districts. ***, **, * show significance at 1%, 5%, and 10%, respectively.

Figure 10: Dynamic Spatial Effects of Trade Liberalization



Note: The figure reports the estimates from the dynamic estimation of the importance of internal trade cost. The dependent variable is log of real sales of industry j in district d. Each point shows the additional change in manufacturing output for a 10% (median) decline in trade cost, in year t relative to 1991, along with the 95% confidence interval. Controls include the industry-district size in 1991 and industry share in district's total output in 1991 interacted with year fixed effects. Additionally, district-industry fixed effects and state-year and industry-year fixed effects are included. The sample drops 11 highly urban and FDI receiving districts. Standard errors are two-way clustered at industry-year and district level.

4.2 Manufacturing Activity: Extensive Margin

In order to study the impact on total number of firms and total employment, we rely on data from the Economic Census (EC) of India. We use data from three rounds of Economic Census 1990, 1998, and 2005. Unlike ASI, which surveys only the formal sector firms, EC surveys all establishments in the country and provides a more complete picture of total firms and workers. The inclusion of informal firms in the analysis is crucial as even though they account for under 20% of the total manufacturing output, they employ over 80% of the manufacturing workforce in India (World Bank Blogs, 2015). Excluding them from the analysis will drastically reduce the actual number of firms and labor employed. The key variables that we use from EC are a firm's district, number of employees, and 4-digit industry code. The estimation strategy is similar to Equation 1. We estimate the following specification:

$$y_{jdt} = \beta_0 + \beta_1 erp_{jt} \times Trade\,Cost_d + \beta_2 erp_{jt} + DMA_{dt} + X_{jd,pre}\delta_t + \lambda_{jd} + \delta_t + \epsilon_{jdt} \tag{2}$$

where y is the outcome of interest that includes (log) number of firms and (log) total labor in industry j district d in year t. A positive sign on β_1 indicates that districts with better international access gained relatively more in terms of number of manufacturing firms and/or manufacturing sector labor. As in Section 4.1, we include year and industry-district fixed effects. $X_{jd,pre}$ contains a district-industry pair's pre-liberalization outcome, y, interacted with the year fixed effects. We weigh the regressions using the 1990 (log) total labor for an industry-district pair. Standard errors are two-way clustered at the 4-digit industry-year and district level.

We present the results for the number of firms and labor separately in Tables 6 and 7, respectively. Column (1) in Table 6 shows the results from Equation 2. In the baseline, we interact industry-district pair's 1990 level of number of firms with year fixed effects to to allow for the fact the growth in the number of firms might respond to the initial level. The coefficient on β_1 is positive and significant at the 1% level. Specifically, for a 10 pp decline in *erp* in an industry, reducing trade cost by 10% leads to an increase in the number of firms by 2.38%. Columns (2) drops 2005 census from the data (pre-GQ). Column (3) drops 11 high FDI inflow districts. Column (4) adds state-year fixed effects and further includes the pre-liberalization level of district-industry output times the year fixed effect as additional control [16]. Column (5) adds to Column(4), the industry-year

[16]This control is from the ASI data. Since the variable fundamentally reflects total economic activity where formal firms account for a very large share, unavailability of informal activity information is not a

fixed effect and removes the state-year fixed effect. The impact of international access stays consistently positive and significant. For the average decline in erp of 98 pp during the sample period, our estimates indicate a relative increase in the number of firms by 23.32% in districts with lower trade cost by the median.

Table 6: International Access and Number of Manufacturing Firms

	Log Number of Firms				
	(1)	(2)	(3)	(4)	(5)
erp	-2.386***	-1.848***	-2.033***	-1.781***	-
	(0.552)	(0.606)	(0.540)	(0.543)	
$erp \times Trade\ Cost$	2.178***	1.755***	1.855***	1.625***	2.613***
	(0.504)	(0.565)	(0.493)	(0.504)	(0.506)
Observations	64,604	41,008	62,348	64,604	64,604
R-squared	0.822	0.910	0.819	0.824	0.860
District-Industry FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	N	N
State-Year FE	N	N	N	Y	N
Industry-Year FE	N	N	N	N	Y

Note: The dependent variable is the log of total number of firms in industry j in district d. erp is the effective rate of protection faced by the industry in year t. *Trade Cost* measures the iceberg trade cost between a district's centroid and its nearest port. Column (2) restricts the period of analysis up to the year 1998 to exclude the effects of the GQ highway. Column (3) drops the 11 districts that received most of the FDI during 1993-2008. Column (4) includes state-year fixed effects and initial district-industry size controls interacted with year fixed effect. Column (5) includes industry-year fixed effects. Standard errors are two-way clustered at the 4-digit industry-year and district level. The sample consists of 98 4-digit industries across 338 districts. ***, **, * show significance at 1%, 5%, and 10%, respectively.

We find very similar patterns for total labor in manufacturing. In the baseline, we interact industry-district pair's 1990 level of total labor with year fixed effects to allow for the fact that growth in labor might be related to the initial labor size. For a 10 pp reduction in *erp*, the point estimates imply an increase in total manufacturing labor employed by 3.9%. Results are robust to the tests performed for total firms as shown in Columns (2) to Column (5) of Table 7. These results show a faster reallocation of labor, within manufacturing, towards sectors with a larger decline in protection, in districts with lower trade costs. Overall, for the average decline in *erp* of 98 pp in our sample period, reduction in trade cost by the median value leads to, on average, a 38.2% increase in total labor in an industry with the average decline in erp, relative to an industry with no change in erp.

major concern.

While EC offers several advantages over ASI in terms of coverage, the one caveat is that we can no longer perform a pre-trend test. EC data is not digitized prior to 1990. Causal interpretations of results in this section should therefore be approached with appropriate caution. However, the inclusion of 1990-level district-industry outcomes interacted with year fixed effects does strengthen the assumption that results in this section are not driven by pre-existing trends in number of firms and total labor.

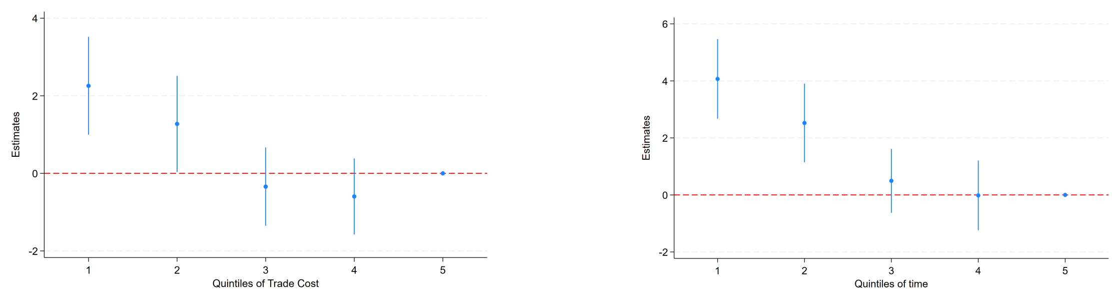
Table 7: International Access and Manufacturing Employment

	Log Employment				
	(1)	(2)	(3)	(4)	(5)
erp	-3.917***	-3.168***	-3.507***	-2.437***	-
	(0.643)	(0.758)	(0.596)	(0.597)	
$erp \times Trade\ Cost$	3.599***	2.992***	3.219***	2.245***	4.137***
	(0.584)	(0.702)	(0.539)	(0.553)	(0.584)
Observations	64,602	41,004	62,346	64,602	64,602
R-squared	0.795	0.880	0.788	0.799	0.829
District-Industry FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	N	N
State-Year FE	N	N	N	Y	N
Industry-Year FE	N	N	N	N	Y

Note: The dependent variable is the log of total labor hired in industry j in district d. erp is the effective rate of protection faced by the industry in year t. *Trade Cost* measures the iceberg trade cost between a district's centroid and its nearest port. Column (2) restricts the period of analysis up to the year 1998 to exclude the effects of the GQ highway. Column (3) drops the 11 districts that received most of the FDI during 1993-2008. Column (4) includes state-year fixed effects and initial district-industry size controls interacted with year fixed effect. Column (5) includes industry-year fixed effects. Standard errors are two-way clustered at the 4-digit industry-year and district level. The sample consists of 98 4-digit industries across 338 districts. ***, **, * show significance at 1%, 5%, and 10%, respectively.

In Figure 11, we show the semi-parametric estimates from the baseline regression in Equation 2 with trade cost divided into quintiles. Additional gains in labor are monotonically increasing in reduced trade costs. This is, however, not the case for number of firms. While districts sufficiently closer to the international gates gain more, the additional gains disappear if the district is sufficiently closer to the farthest quintile. Thus, while gains in output and productivity are monotonic in reduced trade costs, additional gains in number of firms are clustered closer to international gates. This is intuitive, theoretically, as land is a fixed factor and increased demand for land closer to the gates can drive up rents. The benefits of moving closer to gates can outweigh the increased rents up to a certain threshold of trade cost. Beyond the threshold, in districts with better international access as compared to the farthest districts, existing firms still gain in terms of output, but higher rents can hamper new firm entry.

Figure 11: Semi-Parametric Estimates of Trade Cost



(a) Number of Firms

(b) Total Labor

Note: This figure reports the results from a semi-parametric estimation of Equation 2. *Trade Cost* is divided into 5 quintiles. Each point estimate represents the additional effect on the number of firms/labor for the *i*'th quintile relative to the 5th quintile for a 10 pp decline in *erp*.

4.3 Structural Change

Our firm-level analysis provides strong evidence of relatively larger benefits in manufacturing sector activity, from trade liberalization, accruing to regions with better international access. In this section, we examine whether the local labor market responses are also impacted by the proximity to international gates. Using data from multiple rounds of the NSS and the Population Census, we assemble a dataset on district-level shares of labor force employed in agriculture, manufacturing, and services[17] for the years 1987-88, 1991, 1999-2000, and 2004-05. We summarize our key outcome variables in Table 8 below. During this period, there is an overall decline in the share of labor force in agriculture accompanied by an increase in the share of manufacturing and services.

We estimate the following equation:

$$y^k_{dst} = \beta_0 + \beta^k \, Trade\,Cost_d \times Post_t + X^k_{d,pre}\,\delta_t + X_{dt} + \gamma_d + \delta_t + \epsilon_{dt} \tag{3}$$

Here, y^k is the share of labor force in district d, state s and year t employed in sector k, where $k \in \{$*agriculture, manufacturing, services*$\}$. *Trade Cost* measures the iceberg trade cost between district d's centroid and its nearest port. *Post* is a dummy variable for the post-liberalization period and takes the value 1 for 1999-2000 and 2004-05 and 0

[17] We include non-tradeable services like construction in our classification of the service sector.

Table 8: Descriptive Statistics: Sectoral Employment Shares

	Median		Mean	
	Pre	Post	Pre	Post
Agriculture	0.74	0.61	0.68	0.58
Manufacturing	0.07	0.08	0.09	0.11
Services	0.17	0.27	0.20	0.29
#Districts		340		

Note: Pre period is defined as the year 1991. *Post* year refers to the average across 1999-00 and 2004-05. Employment shares are defined as the total employment in sector k as a share of the district's labor force.

for 1991. $X^k_{d,pre}$ is a vector of a district's pre-liberalization characteristics like (log) population, the share of females, primary educated population, working-age population, the share of population residing in urban areas, and average ruggedness[18]. We interact these district-level baseline characteristics with year fixed effects to allow districts differing in terms of initial characteristics to have different effects on sectoral employment shares due to trade liberalization. We also control for the share of labor force employed in sector k in 1991 times year fixed effects to allow for correlations between initial sector size and future employment shares. X_{dt} controls for time-varying district characteristics like domestic market access, annual rainfall, and annual rainfall squared[19]. Rainfall accounts for supply-side shocks in agriculture that could affect sectoral employment shares across districts differentially in any given year. γ_d controls for district fixed effects and δ_t controls for year fixed effects. We cluster standard errors at the district-level and weigh the regressions by the district's 1991 (log) population. The labor force in our sample is comprised of all individuals aged 15-59.

The coefficient of interest β^k measures the causal effect of proximity to an international port on a district's employment share in sector k, on average, post-trade liberalization. $\beta^k < 0$ implies that trade liberalization increases the employment share of sector k as international access improves (reduction in trade cost). Since manufacturing activity increased more in districts closer to international gates than those further away, we expect $\beta < 0$ for manufacturing and $\beta > 0$ for agriculture. Since wages in the manufacturing sector are generally higher than in agriculture, a larger growth in manufacturing employment in districts closer to the ports implies a greater share of people working in

[18]Demographic characteristics are computed from the Population Census of 1991 and data on ruggedness is taken from Asher et al. (2021).

[19]We use Indian Meteorological Department's grid-level data on daily rainfall to compute total rainfall received by district d in year t.

higher-wage sectors. Under the assumption of non-homothetic preferences, the demand for services should rise faster in districts closer to ports leading to, potentially, a greater increase in service employment ($\beta < 0$).

In this setup, β^k is identified by comparing the effects of trade liberalization on districts with similar observable characteristics but different international access. Any threat to identification would arise if factors that affect a district's sectoral employment share are correlated with the district's international access. We identify three potential threats to identification in our setup. First, β^k would be biased if a district's exposure to trade liberalization is correlated with its international access. For instance, if districts near ports systematically have a larger presence of industries that experienced greater tariff cuts during trade liberalization, we might be overestimating the effect of international access for structural change. To eliminate this possibility, we control for a district's tariff exposure based on its industrial composition of 1991 following Topalova (2007)[20]. Other threats include increased FDI in districts with better international access and the launch of the GQ construction in 2001, discussed in detail in Section 4.1. Both channels can lead to upward bias in the estimates. We show that the results are robust to dropping high FDI districts from the sample (this also accounts for any effects driven by highly urban districts as these are also the most urban districts in India) and to restricting the post period to 1999-2000 (pre-GQ) only.

Results: We summarize our results from Equation 3 in Table 9. We begin by examining the results for agriculture in Panel (A). Column (1) contains the estimates from the baseline specification. We find that districts with better international access experience a greater decrease in agriculture employment share due to trade liberalization. Specifically, comparing a district at the median of the trade cost distribution (1.10) to a district with the least trade cost (1.00), we find a 4.24 pp greater decline in agriculture employment share due to trade liberalization at the least trade cost district ($0.424 \times 100 \times (1.10 - 1.00)$). In Columns (2)-(4), we address the threats to identification discussed above. In Column (2), we add a control for districts' tariff exposure from trade liberalization. In Column (3), we drop the 11 districts that received the highest FDI projects during the period of our analysis. In Column (4), we restrict our post period to 1999-00. In Column (5) we add state-year fixed effects to the specification in Column (2) for additional robustness. This is important because states may have varying labor laws and different timelines for implementation, which can affect the reallocation of labor across sectors. Our effect size is stable across specifications and ranges between 3.10 and 4.26 pp for a median reduction in trade cost across districts.

[20]Construction of this measure is discussed in Appendix A.3.

Table 9: International Access and Sectoral Employment Shares

	(1)	(2)	(3)	(4)	(5)
			Panel (A): Agriculture		
$Trade\ Cost \times Post$	0.424***	0.426***	0.404***	0.325**	0.310**
	(0.122)	(0.122)	(0.121)	(0.158)	(0.128)
			Panel (B): Manufacturing		
$Trade\ Cost \times Post$	-0.239***	-0.241***	-0.212***	-0.224***	-0.159***
	(0.059)	(0.059)	(0.056)	(0.072)	(0.060)
			Panel (C): Services		
$Trade\ Cost \times Post$	-0.273***	-0.269***	-0.244**	-0.230*	-0.183*
	(0.100)	(0.101)	(0.098)	(0.125)	(0.106)
Observations	1,019	1,019	986	678	1,010
District FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	N
State-Year FE	N	N	N	N	Y

Note: The dependent variable is the share of the labor force in a district employed in sector k. *Trade Cost* measures the iceberg trade cost between a district's centroid and its nearest port. *Post* is an indicator variable equal to 1 if the year is 1999-00 or 2004-05 and 0 if the year is 1991. Column (2) controls for district-level exposure to tariffs in year t. Column (3) drops the 11 districts that received most of the FDI (and most urban) during 1993-2008. Column (4) restricts the post-period to 1999-00 to exclude the effects of the GQ highway. Column (5) includes state-year fixed effects. The sample consists of 340 districts. Standard errors are clustered at the district level. ***, **, * show significance at 1%, 5%, and 10%, respectively.

Panel (B) summarizes the results for the manufacturing sector. The point estimates of β are negative and highly significant. In Column (1), a median reduction in trade cost across districts leads to a 2.39 pp increase in the share of manufacturing employment on average due to trade liberalization. Consistent with our previous estimates, stronger gains in manufacturing activity in districts closer to international gateways indeed translate into a greater increase in the manufacturing employment share in these districts. Considering all specifications, the effect size ranges between 1.59 to 2.41 pp for a median reduction in trade cost across districts. These effects are smaller than the decline in agriculture shares which indicates some reallocation towards services.

In Panel C, we examine the service sector. The baseline point estimate shows a significant increase of 2.73 pp in service share of employment due to trade liberalization for a median reduction in trade cost across districts. This is in line with the non-homothetic preferences hypothesis that higher incomes lead to increased demand for services. The effects for services fall within the range of 1.83 to 2.73 pp. Overall, we find evidence for a strong positive role of international access in facilitating structural change, due to a large trade liberalization episode. To confirm that these effects are not driven by pre-trends in employment shares, we conduct a placebo test using data from pre-liberalization periods of 1987-88 and 1991. We estimate Equation 3 by assigning 1991 as the treatment year ($Post = 1$) and 1987-88 as the control year ($Post = 0$). Results in Table 10 indicate that coefficients are in the opposite direction and not statistically significant for all three sectors.

Table 10: Placebo Test: International Access and Sectoral Employment Shares

	Agriculture	Manufacturing	Services
	(1)	(2)	(3)
$Trade\ Cost \times Post'$	-0.121	0.055	0.125
	(0.206)	(0.085)	(0.148)
Observations	674	674	674
District FE	Y	Y	Y
State-Year FE	Y	Y	Y

Note: The dependent variable is the share of the labor force in a district employed in sector k. *Trade Cost* measures the iceberg trade cost between a district's centroid and its nearest port. *Post′* is an indicator variable equal to 1 if the year is 1991 and 0 if the year is 1987-88. The sample consists of 337 districts. Standard errors are clustered at the district level. ***, **, * show significance at 1%, 5%, and 10%, respectively.

Benchmarking the importance of international access: In order to understand the economic importance of our estimates for structural transformation, we conduct a simple exercise. We categorize each district as below or above-median based on its trade cost. We re-run specification (3) but with travel time now divided into 2 groups as,

$$y^k_{dst} = \beta_0 + \beta^k Below\ Median_d \times Post_t + X^k_{d,pre}\,\delta_t + X_{dt} + \gamma_d + \eta_{st} + \epsilon_{dt} \tag{4}$$

where $Below\ Median$ takes the value 1 if the district is below the median in the distribution of trade costs, and 0 otherwise. In essence, we are running the specification with all the pre-liberalization controls, geographic controls, tariff exposure of districts, domestic market access, and state-year fixed effects. The estimates of β^k give us the average additional change in employment share in sector k, in below-median trade cost districts relative to the above median, due to trade liberalization. We use these estimates to examine how much of the overall additional change in the below-median districts is explained by the proximity to international gates. Throughout the exercise, we try to be conservative in our analysis. Apart from using the specification with all the controls discussed above, we further use the lower bound from the 95% confidence intervals (CI) of the point estimates of β_k. The results are presented in Table 11.

Table 11: Benchmarking Exercise for Structural Change Estimates

	Agriculture	Manufacturing	Services
	(1)	(2)	(3)
$Below\ Median \times Post$	-0.032***	0.013**	0.020**
	(0.010)	(0.005)	(0.009)
Observations	1,010	1,010	1,010
95% C.I.	[-0.053, -0.011]	[0.003, 0.023]	[0.003, 0.037]
District FE	Y	Y	Y
State-Year FE	Y	Y	Y

Note: The dependent variable is the share of the labor force in a district employed in sector k. *Below Median* is an indicator variable for a district with below median *Trade Cost*. *Post* is an indicator variable equal to 1 if the year is 1990-00 or 2004-05 and 0 if the year is 1991. The sample consists of 337 districts. Standard errors are clustered at the district level. ***, **, * show significance at 1%, 5%, and 10%, respectively.

As expected, the additional effects on the share of labor employed in agriculture are negative and significant, and the additional effects on the share of labor in manufacturing and services are positive and significant. In data, the overall additional decline in the share of labor force in agriculture from the pre- to the post-period in the below-median districts relative to the above-median districts is 0.96 pp. The lower bound of the 95% CI for agriculture, 1.1 pp explains 120% of the overall additional decline in the below-median districts. This corresponds to 2.02 million people and thus, better international access can explain, at least, an additional 2.02 million people leaving agriculture in the below-median districts. Similarly, the lower bound of the 95% CI for manufacturing explains 76% of the observed additional increase in the share of labor force in the manufacturing sector. This translates to an additional 0.5 million people. Thus, better international access can explain over a million additional jobs created in the manufacturing sector in districts with below-median trade cost. In the service sector, the lower bound explains roughly 16% of the additional increase in the share of labor force in services in the below-median districts. This corresponds to roughly 0.45 million additional service jobs in the below-median districts due to better international access.

4.4 Internal Migration

Migration is an important channel for reducing spatial disparities in income and opportunities. India is characterized by a low rate of inter-district migration. As per the 2011 census, the bulk of the movement is within a district (62%). Inter-district movement within a state is a further 26% and only 12% of the movement is inter-state (World Bank Blog, 2019)[21]. We have established that districts more integrated with world markets create more opportunities in the modern sectors following India's trade liberalization. Using the 2001 census data, we test whether, post liberalization, the migration inflow is larger into districts with lower trade costs and outflows are larger from districts with higher trade costs. We estimate the following equation using the Poisson Pseudo Maximum Likelihood procedure (PPML):

$$\begin{aligned} MigFlow_{ds,os'} = \beta_0 + \beta_1 Trade\,Cost_d + \beta_2 Trade\,Cost_o + \beta_3 Travel\,Time_{do} \\ + \alpha_1 L_{do} + X_d + X_o + \eta_{ss'} + \epsilon_{ds,os'} \end{aligned} \tag{5}$$

where $MigFlow$ is the stock of migrants in destination district d in state s from origin district o in state s'. $TradeCost_d$ and $TradeCost_o$ are the trade costs of destination and origin districts, respectively. $Travel\,Time_{do}$ is the travel time (in hours) between the districts and L_{do} is the probability that people in districts d and o speak the same mother tongue, and thus can communicate more easily. X contains pre-liberalization characteristics like (log) population and (log) mean night light intensity of destination and origin districts. Night light intensity is a proxy for district's GDP and controls for the fact that migration

[21] These findings have been confirmed by Ghose (2021) who shows that migration is reducing in distance and state borders are important barriers.

inflows (outflows) might be larger in more (less) developed districts. $\eta_{ss'}$ is the destination state-origin state fixed effect that accounts for any state-to-state specific migration costs. We restrict the sample to people who migrated in the last 10 years (1991-2001) to avoid commenting on any pre-liberalization flow patterns. Standard errors are clustered at the destination state-origin state level.

Our coefficients of interest are β_1 and β_2. A negative sign on β_1 shows that controlling for the bilateral distance, common language, population, GDP, and state pair fixed effects, the inflow of migrants is higher in districts with low trade costs. Similarly, we expect a positive coefficient on β_2 indicating a higher outflow from districts less integrated with world markets. Results are shown in Table 12. We obtain a negative and significant estimate of β_1 and a positive and significant estimate of β_2, confirming our hypothesis. Thus, migration inflows (outflows) are larger for districts that gain more (less) from trade liberalization. As expected, the coefficient on bilateral travel time is negative, and common language share is positive.

Table 12: Migration between districts post liberalization

	Migrants
	(1)
$Trade\,Cost^d$	-7.611***
	(2.142)
$Trade\,Cost^o$	5.332**
	(2.210)
$Travel\,Time_{do}$	-0.151***
	(0.020)
L_{do}	0.236
	(0.345)
Observations	118,523
Dest-Orig State FE	Y

Note: The independent variable is the stock of migrants. Controls include (log) population of 1991 and (log) mean night light intensity of 1994 (first available data point) of destination and origin districts. Standard errors are clustered at the destination state-origin state level. ***, **, * show significance at 1%, 5%, and 10%, respectively.

We, therefore, document a new fact about migration in India. While migration flows across districts might be low, the post-trade liberalization inflows are larger in districts with low trade costs and outflows are larger from districts with high trade costs.

5 Conclusion

This paper examines a large-scale trade reform to highlight the crucial role of internal geography for regional structural transformation, in the presence of high internal trade costs. Focusing on India's 1990s trade liberalization, our findings indicate that regions with better access to international markets, via international ports, experienced significantly stronger growth in the manufacturing sector. These uneven gains facilitated a greater reallocation of labor from agriculture to manufacturing and services in regions with better international access. Notably, these differential effects persist over a long period, spanning up to 15 years following trade liberalization. In future work, we plan to estimate the implications of internal trade costs on welfare and regional inequality through the lens of a quantitative spatial model.

References

Alder, Simon, "Chinese roads in India: The effect of transport infrastructure on economic development," *Available at SSRN 2856050*, 2016.

Allen, Treb and David Atkin, "Volatility and the Gains from Trade," *Econometrica*, 2022, *90* (5), 2053–2092.

Asher, Sam and Paul Novosad, "Rural roads and local economic development," *American economic review*, 2020, *110* (3), 797–823.

_, Tobias Lunt, Ryu Matsuura, and Paul Novosad, "Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform," *The World Bank Economic Review*, 2021, *35* (4).

Asturias, Jose, Manuel García-Santana, and Roberto Ramos, "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 2019, *17* (6), 1881–1940.

Barnwal, Prabhat, Jonathan I Dingel, Daniil Iurchenko, Pravin Krishna, and Eva Van Leemput, "Internal Trade Barriers in India," 2024.

Baum-Snow, Nathaniel, J Vernon Henderson, Matthew A Turner, Qinghua Zhang, and Loren Brandt, "Does investment in national highways help or hurt hinterland city growth?," *Journal of Urban Economics*, 2020, *115*, 103124.

Belmar, José, "Trade and Structural Change: Evidence from Colombia and The Panama Canal," *Unpublished Manuscript*, 2023. Available at: https://www.josebelmar.com/research#h.n7gzxdrdiofo.

Brooks, Wyatt and Kevin Donovan, "Eliminating uncertainty in market access: The impact of new bridges in rural Nicaragua," *Econometrica*, 2020, *88* (5), 1965–1997.

Castro-Vincenzi, Juanma and Benny Kleinman, "Intermediate input prices and the labor share," *Princeton University, unpublished manuscript, https://www. castrovincenzi.com/research/blog-post-title-two-49acg (Accessed December 2 2023)*, 2022.

Cheung, Terry and Han Yang, "Transportation Networks, Technology Adoption, and Structural Transformation," *Technology Adoption, and Structural Transformation*, 2024.

Corden, Warner Max, "The structure of a tariff system and the effective protective rate," *Journal of Political Economy*, 1966, *74* (3), 221–237.

Coşar, A Kerem and Pablo D Fajgelbaum, "Internal geography, international trade, and regional specialization," *American Economic Journal: Microeconomics*, 2016, *8* (1), 24–56.

Cravino, Javier and Sebastian Sotelo, "Trade-induced structural change and the skill premium," *American Economic Journal: Macroeconomics*, 2019, *11* (3), 289–326.

Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

_ **and Richard Hornbeck**, "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Dreze, Jean and Christian Oldiges, "Work in Progress," in "Frontline," Vol. 26 2009, pp. 101–105.

Fajgelbaum, Pablo and Stephen J Redding, "Trade, structural transformation, and development: Evidence from Argentina 1869–1914," *Journal of political economy*, 2022, *130* (5), 1249–1318.

Fan, Tianyu, Michael Peters, and Fabrizio Zilibotti, "Growing like India—the unequal effects of service-led growth," *Econometrica*, 2023, *91* (4), 1457–1494.

Gallé, Johannes, Daniel Overbeck, Nadine Riedel, and Tobias Seidel, "Place-based policies, structural change and female labor: Evidence from India's Special Economic Zones," *Available at SSRN 4210059*, 2022.

Ghani, Ejaz, Arti Grover Goswami, and William R Kerr, "Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing," *The Economic Journal*, 2016, *126* (591), 317–357.

Ghose, Devaki, *Trade, Internal Migration, and Human Capital: Who Gains from India's IT Boom?*, World Bank, 2021.

Goldberg, Pinelopi K, Amit K Khandelwal, Nina Pavcnik, and Petia Topalova, "Multiproduct firms and product turnover in the developing world: Evidence from India," *The Review of Economics and Statistics*, 2010, *92* (4), 1042–1049.

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova, "Imported intermediate inputs and domestic product growth: Evidence from India," *The Quarterly journal of economics*, 2010, *125* (4), 1727–1767.

Goyal, Surinder K, "Political economy of India's economic reforms," *Institute for Studies in Industrial Development (ISID) Working Paper*, 1996, *4.*

Hasan, Rana, Devashish Mitra, and Krishnarajapet V Ramaswamy, "Trade reforms, labor regulations, and labor-demand elasticities: Empirical evidence from India," *The Review of Economics and Statistics*, 2007, *89* (3), 466–481.

Hyun, Yeseul, Shree Ravi et al., "Place-based development: evidence from special economic zones in India," *Institute for Economic Development (IED) Working Paper Series, Boston University*, 2018.

ILO, "India wage report: Wage policies for decent work and inclusive growth," 2018.

Leemput, Eva Van, "A passage to India: Quantifying internal and external barriers to trade," *Journal of International Economics*, 2021, *131*, 103473.

Melitz, Marc J, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *econometrica*, 2003, *71* (6), 1695–1725.

Mohan, Rakesh et al., "The India infrastructure report: policy imperatives for growth and welfare," *(No Title)*, 1996.

Peter, Alessandra and Cian Ruane, "The aggregate importance of intermediate input substitutability," Technical Report, National Bureau of Economic Research 2023.

Sahu, Prasanta K, Agnivesh Pani, and Georgina Santos, "Freight traffic impacts and logistics inefficiencies in India: Policy interventions and solution concepts for sustainable city logistics," *Transportation in Developing Economies*, 2022, *8* (2), 31.

Sukhtankar, Sandip, "India's national rural employment guarantee scheme: What do we really know about the world's largest workfare program?," in "Brookings-NCAER India Policy Forum," Vol. 13 2017, pp. 231–286.

Topalova, Petia, "Trade liberalization, poverty and inequality: Evidence from Indian districts," in "Globalization and poverty," University of Chicago Press, 2007, pp. 291–336.

_ **and Amit Khandelwal**, "Trade liberalization and firm productivity: The case of India," *Review of economics and statistics*, 2011, *93* (3), 995–1009.

Varshney, Ashutosh, "Mass politics or elite politics? India's economic reforms in comparative perspective," *The Journal of Policy Reform*, 1998, *2* (4), 301–335.

Vogel, Kathryn Baragwanath, Gordon H Hanson, Amit Khandelwal, Chen Liu, and Hogeun Park, "Using Satellite Imagery to Detect the Impacts of New Highways: An Application to India," Technical Report, National Bureau of Economic Research 2024.

Xu, Yang and Xi Yang, "Access to ports and the welfare gains from domestic transportation infrastructure," *Journal of Urban Economics*, 2021, *126*, 103392.

A Data Appendix

A.1 Methodology to obtain district identifiers in ASI data

This section provides details of the matching exercise undertaken to augment district identifiers to the ASI data from 1987-88 to 2009-10. We begin by collecting all available publications of MoSPI on district codebooks for the period of our ASI sample. This includes the district codebooks released for ASI 2009-10, the Economic Census (EC) of 2005, 1998, and 1990, and the National Sample Surveys (NSS) of 2004-05, 1999-2000, 1993-94 and 1987-88.

The first step involves identifying which publications out of the above-mentioned can be used as a valid codebook for the ASI rounds in our sample. To do this, we perform string matching algorithms over successive codebooks going backwards in time, i.e., we match ASI 2009-10 to EC 2005, EC 2005 to NSS 2004-05, and so on. This helps in identifying the years in which MoSPI changed the codebooks. We find that codebooks are identical for ASI 2009-10, EC 2005, and NSS 2004-05. The district nomenclature changes in NSS 1999-2000 and is identical across EC 1998 (except for the 3 states) and NSS 1993-94. Next, changes occur for EC 1990 and NSS 1987-88. To sum up, district nomenclatures in ASI are expected to be different for the periods 1987-90, 1990-93, 1993-2000, and 2000-04, 2004-10. This is further strengthened by examining the changes in the state codebooks available for each round of ASI. Inspection of ASI state codebooks reveals that state codes were similar from 1987-90, 1990-2000, and 2000-10. Combining the two, we obtain the following bands: (i) ASI 2009-10 codebook applies to ASI 2000-10; (ii) a combined codebook of NSS 1999-00 and EC 1998 applies to ASI 1993-2000; (iii) EC 1990 codebook applies to ASI 1990-93; and (iv) NSS 1987-88 codebook applies ASI 1987-90.

Next, as district boundaries change over time, we need to create a concordance across these codebooks so that district boundaries are consistent over the period 1987-2010. To do so, we use the codebook of EC 1990 as the basefile and construct a concordance of all district codebooks from 1990-2009 with district EC 1990. We use a string-matching algorithm that matches two codebooks on common district names. We manually inspect and match all districts that are matched with a similarity score ¡ 90%. As district boundaries change over time, to ensure that the district codes in the concordance consistently represent the district boundaries that existed in 1990, we undertake the following process:

- If a new district was carved out of one parent district, we map the new district to the code of the EC 1990 parent district. For example, in ASI 2009-10 [2723] Mumbai Suburban was carved out of the EC 1990 district [1401] Greater Mumbai. We map

[2723] Mumbai Suburban in ASI 2009-10 to [1401] Greater Mumbai in EC 1990.

- If a new district is carved out of multiple parent districts, we aggregate all parent districts and assign them the code of the parent with the smallest code in EC 1990. For example, in ASI 2009-10 [0318] Mohali was formed by combining parts of EC 1990 districts [2008] Roopnagar and [2009] Patiala. We aggregate Roopnagar and Patiala in EC 1990 and code the aggregated district as [2008]. We map [0318] Mohali in ASI 2009-10 to [2008] aggregated Roopnagar in EC 1990.

Lastly, we map the codebooks from 1987-90 to EC 1990 such that the district boundaries are representative of the 1987 level. This exercise yields 345 unique districts at 1987 administrative boundaries. We then apply these concordances to the ASI rounds so that each district in an ASI round gets mapped to a district in 1987. As a result, we are able to consistently identify districts across all rounds of ASI from 1987-88 to 2009-10.

A.2 Domestic Market Access

Following Donaldson and Hornbeck (2016), we compute a first-order approximation of domestic market access of a district as follows:

$$DMA_{dt} = \sum_{d' \neq d} Trade\,Cost_{dd'}^{-\theta} N_{d't}$$

where, d is the origin district and d' is the destination district. *Trade Cost* is the iceberg trade cost of shipping a good from d to d' and N is the population of the destination district d'. As annual population data is not available, we use district populations from the 1991 and 2001 Population Censuses and derive the compounding annual growth rate to predict populations for 1987-2005. θ is the trade elasticity which is assumed to be 1.5 following Vogel et al. (2024). During the period of our analysis, there are no major changes to the Indian road network, so we estimate trade costs using the road network of 1996. Thus, the only source of variation in *DMA* comes from population changes. In all specifications, we control for the log of *DMA*.

A.3 District-level Trade Liberalization Exposure

Following Topalova (2007), we create a measure of a district d's exposure to trade liberalization in year t as follows:

$$Exposure_{dt} = \sum_j \omega_{jd,1991} \times \tau_{jt}$$

Here, $\omega_{jd,1991}$ is the share of manufacturing labor in district d employed in industry j in the year 1991 (i.e., pre-liberalization).τ_{jt} is the tariff on industry j in year t.